Confidence bands for Brownian motion and applications to Monte Carlo simulation

W.S. Kendall*, J.-M. Marin† and C.P. Robert‡

July 4, 2006

Abstract

Minimal area regions are constructed for Brownian paths and perturbed Brownian paths. While the theoretical optimal region cannot be obtained in closed form, we provide practical confidence regions based on numerical approximations and local time arguments. These regions are used to provide informal convergence assessments for both Monte Carlo and Markov Chain Monte Carlo experiments, via the Brownian asymptotic approximation of cumulative sums.

Keywords: Boundary crossing probability, Brownian motion, Central Limit Theorem, CUSUM, local time, Monte Carlo path, simultaneous confidence region.

1 Introduction

A problem that has not been addressed so far in the literature is the construction of a confidence band on a Brownian motion $(W_t)_{t\in[0,1]}$. A *confidence band* on a random function amounts to the derivation of a function u on $[0,1]$ such that for $0 < \alpha < 1$)

$$\mathbb{P}(-u(t) \le W(t) \le u(t), t \in [0,1]) = 1 - \alpha\,.$$

In fact, the computation of the probability that a standard Brownian motion remains between two given boundaries is a complex question that has occupied many researchers over many years (see, e.g., Anderson, 1960; Robbins and Siegmund, 1970; Durbin, 1971; Lerche, 1986; Durbin, 1992; Daniels, 1996; Borodin and Salminen, 2002; Li, 2003). In this literature, explicit representations of two-sided boundary crossing probabilities are extremely rare and mostly address linear cases (Anderson, 1960; Hall, 1997), though explicit computations are available for a few nonlinear boundaries (Robbins and Siegmund, 1970; Daniels, 1996; Novikov et al., 1999). For most boundary functions computation of two-sided boundary crossing probabilities for the Brownian motion requires some level of approximation.

Our goal is to optimise the choice of the bound u against an area criterion (using as analogy the construction of the standard confidence interval to minimize the length for a given coverage); we therefore consider the optimisation problem

$$\arg\min_{u\ge 0} \int_0^1 u(t)\mathrm{d}t\,.$$

This minimization problem does not have a closed form solution and it is thus extremely difficult to solve. We will produce in Section 2 an approximation that is based on numerical approximations and local time arguments.

While the problem of deriving such an optimal band on the Brownian motion is of interest *per se*, it also has important bearings in various fields that take advantage of Brownian approximations. Obvious examples are the areas of sequential testing (Robbins and Siegmund, 1970) and of nonparametric tests (Hollander and Wolfe, 1999). We choose to focus here on a third area where the derivation of such bands has importance, namely the design and evaluation of Monte Carlo methods, as explained in Section 3. This particular application also motivates the search for an optimal band on a Brownian motion with random drift, $(W_t + t\varepsilon)_{t\in[0,1]}$, that appears naturally in Monte Carlo setups, as detailed in Section 4. The corresponding optimal band is constructed in Section 5. Section 6 illustrates the confidence procedures in a specific Monte Carlo setting and Section 7 concludes the paper.

*Department of Statistics, University of Warwick, Coventry

† *Corresponding author:* Project Select, INRIA Futurs, Université d'Orsay, 91425 Orsay, France

‡CEREMADE, Université Paris Dauphine and CREST, INSEE, Paris

2 Optimal confidence bands for a Brownian motion

Let $(W_t)_{t\in[0,1]}$ be a standard Brownian motion (Feller, 1971). A generic confidence band on $(W_t)_{t\in[0,1]}$ is a random set of the form

$$\{\omega;\, -v(t) \le \omega(t) \le u(t)\,,\, t\in[0,1]\}$$

associated with the upper and lower bounds, $u(t) \ge 0$ and $v(t) \ge 0$, $(t \in [0,1])$. A natural measure of tightness of this band is its area $\int_0^1 (u(t)+v(t))\mathrm{d}t$ and an optimal band is defined as minimising this area over all possible bounds (u, v) under the constraint that

$$\mathbb{P}(-v(t) \le W(t) \le u(t), t\in[0,1]) = 1-\alpha\,.$$

From the symmetry of the standard Brownian motion, if the minimum is reached and is unique then we have $v = u$ and we can therefore replace the above problem with

$$\min_u \left(\int_0^1 u(t)\mathrm{d}t\right) \qquad (1)$$
$$\text{subject to}$$
$$\mathbb{P}\left\{-u(t) \le W(t) \le u(t)\,,\, t\in(0,1]\right\} = 1-\alpha\,.$$

2.1 Local time solution

Because of the intractable nature of the derivation of the confidence band, we now consider a related problem based on local times.

Starting with the original problem (1), we can consider the dual problem

$$\min_u \left\{1-\mathbb{P}\left(-u(t) \le W(t) \le u(t)\,,\, t\in(0,1]\right)\right\}$$
$$\text{subject to } \int_0^1 u(t)\mathrm{d}t = \beta\,, \qquad (2)$$

where $\beta > 0$ is a given constant. This formulation suggests the study of an alternative and easier problem, in which we replace the probability by a local time expectation:

$$\min_u \mathbb{E}\left[L^u(1) + L^{-u}(1)\right]$$
$$\text{subject to } \int_0^1 u(t)\mathrm{d}t = \beta\,. \qquad (3)$$

Here, $L^u(s)$ is the *local time* accumulated by W along the curve u up to time s, to be defined below, and we require u to be C^1-smooth.

This new problem is easier because local time is additive. Therefore, firstly, we may restrict attention to the one-sided problem, that is, replace $\mathbb{E}\left[L^u(1) + L^{-u}(1)\right]$ by $\mathbb{E}\left[L^u(1)\right]$ and, secondly, it is then possible to localize the minimization problem in time, as now detailed.

Definitions of *local time* for Wiener processes and for other semimartingales can be found in Revuz and Yor (1999, Ch. VI); for our purposes we use the Itô-Tanaka formula (Revuz and Yor, 1999, Ch. VI, Theorem 1.2) which defines a local time L for a continuous semimartingale Y, started at 0, with increasing process $\langle Y, Y\rangle(t) = t$. (These conditions are equivalent to the requirement that Y be a Brownian motion with possibly non-stationary drift.) Applying the Itô-Tanaka formula to $W-u$, we obtain the following formula for the local time L^u of W at the time-varying level u (this is where the C^1-smooth condition on u is required):

$$\frac{1}{2}L^u(t) = \int_0^t u'(s)\mathbb{I}_{[W(s)>u(s)]}\mathrm{d}s + \int_0^t \mathbb{I}_{[W(s)\le u(s)]}\mathrm{d}W(s) - \min\{W(t), u(t)\}\,.$$

The local time $L^u(t)$ can be also interpreted as the time spent by W close to the curve u up to time t (see Revuz and Yor, 1999, Ch. VI, Corollary 1.9);

$$L^u(t) \quad = \quad \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^t \mathbb{I}_{[u(s)\le W(s)<u(s)+\varepsilon]}\mathrm{d}s\,.$$

Therefore, as long as W has a continuous transition probability density $p_{s,t}(x,y)$ (the density of $W(t)$ given $W(s) = x$), we can write

$$\mathbb{E}[L^u(1)] \quad = \quad \int_0^1 p_{0,s}(0,u(s))\mathrm{d}s\,.$$

In our framework, $W(s)$ is normally distributed with mean zero and variance $\tau(s)$. Hence the resolution of the minimization problem (3) amounts to solving

$$\min_u \left(\frac{1}{\sqrt{2\pi}}\int_0^1 \exp\left(-\frac{u(s)^2}{2\,\tau(s)}\right)\frac{\mathrm{d}s}{\sqrt{\tau(s)}}\right)$$
$$\text{subject to } \int_0^1 u(t)\mathrm{d}t = \beta\,.$$

This minimization problem localizes: using a variation $u(s)+\varepsilon h(s)$ and differentiating with respect to

ε, we find the condition for extremality to be

$$\frac{u(s)}{\sqrt{\tau(s)}} \exp\left(-\frac{1}{2}\left(\frac{u(s)}{\sqrt{\tau(s)}}\right)^2\right) \quad = \quad \kappa\tau(s)\,,$$

where κ is the variational constant (or Lagrange multiplier) connected to β. (In fact the problem localizes; this condition holds in an almost-everywhere sense even if we suppose u to be merely measurable.) There are no solutions to this equation if $\kappa\tau(s) > 1/\sqrt{e}$, a unique solution if equality holds, and two solutions (one above $\sqrt{\tau(s)}$, one below) if $\kappa\tau(s) < 1/\sqrt{e}$. A second variation argument shows that the minimization problem is solved by the larger solution if any solution exists, and otherwise we should take $u = 0$ (when $\kappa\tau(s) > 1/\sqrt{e}$).

Taking $\psi(a)$ to be the larger solution to $\psi e^{-\psi^2/2} = a$, if solutions exist, we see that the required minimizing u is given by

$$u^\star(s) \quad = \quad \begin{cases} \psi(\kappa\tau(s))\sqrt{\tau(s)} & \text{if } \kappa\tau(s) \leq 1/\sqrt{e}\,, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The function ψ can be expressed in terms of the so-called `ProductLog` or Lambert W function, the analytic solution $w(z)$ of $we^w = z$ (Corless et al., 1996). Indeed, selecting an appropriate branch of `ProductLog` (using the conventions of *Mathematica*), we find

$$\psi(a) \quad = \quad \sqrt{-\texttt{ProductLog}(-1, -a^2)} \text{ for } a \leq 1/\sqrt{e}\,.$$

The `gsl` package of R contains the function `lambert_Wm1` and

$$\psi(a) \quad = \quad \sqrt{-\texttt{lambert\_Wm1}(-a^2)} \text{ for } a \leq 1/\sqrt{e}\,.$$

Since W is the Brownian motion, $\tau(s) = s$ and the optimal solution $u^\star$ can be easily computed either via *Mathematica* or R. For this type of function the resolution of the original problem then amounts to the computation of the appropriate constant κ to achieve a proper coverage level. For $\alpha = 0.05$, our simulations led to $\kappa = 0.105$ (see below for details).

2.2 Numerical results

We denote by $P_1(u)$ the coverage probability $\mathbb{P}\{|W(t)| \leq u(t), t \in [0,1]\}$. Since the analytical resolution of the optimisation problem for the coverage probability seems to be completely out of reach, we consider a numerical approach based on a partition $t_0 = 0 < t_1 < t_2 < \ldots < t_n = 1$ of the interval $[0,1]$ of size $n \geq 1$, with $\delta t_i = t_i - t_{i-1}$ and $\beta_i = u(t_i)$. We first review some approximation techniques found in the literature. We will distinguish below between three types of approximation, though there exist many other possible approaches in the literature, such as polynomial and Poisson approximations, and Girsanov transformations.

The first type of approximation is based on the approach of Wang and Potzelberger (1997, 2001). These authors show that

$$P_1(u) = \lim_{n \longrightarrow \infty} \mathbb{E}[\ell(W(t_1), W(t_2), \ldots, \ldots, W(t_n)]$$

for any continuous function u, where $\ell(x_1, \ldots, x_n)$ is defined by

$$\prod_{i=1}^{n} \mathbb{I}_{(-\beta_i < x_i < \beta_i)} \left[1 - \exp\left[-\frac{2}{\delta t_i}(-\beta_{i-1} - x_{i-1}) \times (-\beta_i - x_i)\right] - \exp\left[-\frac{2}{\delta t_i}(\beta_{i-1} - x_{i-1})(\beta_i - x_i)\right]\right].$$

Using this result, Potzelberger and Wang (2001) derive the following Monte Carlo estimator of $P_1(u)$

$$\widehat{P_1}(u)_{n,N} = \frac{1}{N}\sum_{j=1}^{N} \ell(W_j(t_1), W_j(t_2), \ldots, W_j(t_n))$$

where $W_1, \ldots, W_N$ are N independent standard Brownian motions on $[0,1]$. As n and N go to infinity, $\hat{P}_1(u)_{n,N}$ converges almost surely to $P_1(u)$. Under the assumptions that the boundary u is twice continuous differentiable with $u''(0) \neq 0$ and $u''(t) = 0$ at most in finitely many points $t \in]0,1]$, the authors proposed a special rule for choosing a sequence of "optimal partitions". Note that it is also possible to use a crude Monte Carlo method to approximate $P_1(u)$, that is to estimate $P_1(u)$ with

$$\widetilde{P_1}(u)_{n,N} = \frac{1}{N}\sum_{j=1}^{N}\prod_{i=0}^{n} \mathbb{I}_{\{-\beta_i \leq W_j(t_i) \leq \beta_i\}}.$$

However, the variance of the Monte Carlo estimator of Potzelberger and Wang (2001) is inevitably smaller than the one of a crude Monte Carlo method insofar as $\ell(x_1, \ldots, x_n) \leq 1$ for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

The second type of approximation is due to Novikov et al. (1999), based on the representation

$$P_1(u) = \mathbb{E}\left[\prod_{i=0}^{n-1} p_i\left\{u | W(t_i), W(t_{i+1})\right\}\right],$$

Class	$u(t)$	minimum	$\int_0^1 u(t)\mathrm{d}t$
1	a	$a = 2.242$	2.242
2	$a + bt$	$a = 1,\ b = 1.77$	1.885
3	$a + b\sqrt{t}$	$a = 0.3,\ b = 2.35$	1.866
$\star$	$u^\star(t)$	$\kappa = 0.105$	1.865

Table 1: Optimal solutions for three classes of parametrised boundaries, along with the local time solution u^* (the two last lines are based on $5,000$ simulated Brownian paths with partition size 500).

where $p_i(u|x,y) = \mathbb{P}(-u(t) < W(t) < u(t), t_i \leq t \leq t_{i+1}|W(t_i) = x, W(t_{i+1}) = y)$. Using u_n, a piecewise linear approximation of u with connecting points $(t_i, u(t_i)),\quad i = 0,\ldots,n$, Hall (1997) calculated the conditional distribution of crossing the upper and lower linear boundaries. Using this result, Novikov et al. (1999) built a fast algorithm that calculates $P_1(u_n)$. Finally, as $u_n(t)$ goes to $u(t)$ uniformly on $[0,1]$, it follows from the continuity property of probability measures that $\lim_{n\longrightarrow\infty} P_1(u_n) = P_1(u)$. Therefore, as n tends to infinity, this approximation gives a convergent approximation of $P_1(u)$.

For the third type of approximation, Durbin (1971), Park and Schuurmann (1976) and Loader and Deely (1987) have proposed the use of certain Volterra integral equations related to the distribution of the first exit time from a region defined by two given boundaries. Based on the discretization of those integral equations, Loader and Deely (1987) have given a numerical solution that is very efficient and we will use this approximation in our numerical experiments.

In those experiments we have only considered three classes of functions besides the solution $u^\star(t)$ given by the local time approximation: the constant class $u(t) = a$ $(a > 0)$, the linear class $u(t) = a + bt$ $(a, b > 0)$, and the square root class $u(t) = a + b\sqrt{t}$ $(a, b > 0)$. For each class, we have derived parameters such that $P_1(u) = 0.95$ and $\int_0^1 u(t)\mathrm{d}t$ is minimum. The first and second classes allow for analytical expressions for boundary crossing probabilities while the third class requires the approximation method of Loader and Deely (1987). Table 1 presents the results of this numerical experiment. Figure 1 shows 5000 simulated Brownian paths together with the solutions in each class.

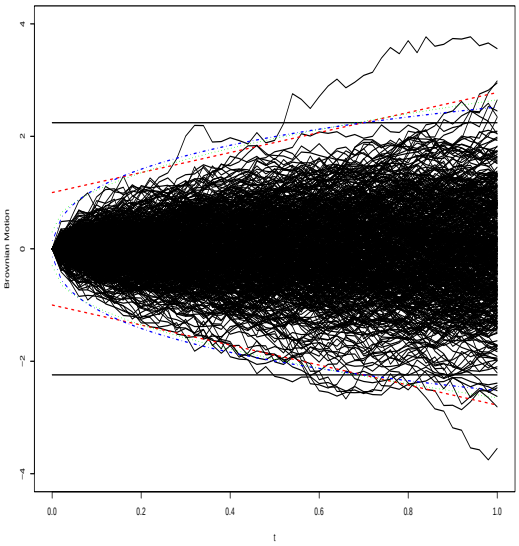

Figure 1: 5000 Brownian paths and functions u such that $P_1(u) = 0.95$, as given by Table 1. The two last functions in Table 1 are almost identical.

As clear from both Figure 1 and Table 1, the gain reached in increasing the complexity of u is very marginal: while the local time approximation function $u^\star$ indeed does better, it is striking and reassuring that the difference is infinitesimal. We also ran simulations for other classes of functions, such as $u(t) = a + bt + ct^2 + d\log(t)$, with no visible improvement. We thus propose to use either $u^\star$ or the simpler function $u(t) = 0.3 + 2.35\sqrt{t}$ for constructing a confidence region of level 0.95.

3 Monte Carlo applications

3.1 Monte Carlo confidence interval

The Monte Carlo method (see, e.g., Robert and Casella, 2004) is validated by both the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) in that (a) the (almost sure) convergence of an average like

$$\widehat{\mathfrak{I}}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \tag{5}$$

to

$$\mathfrak{I} = \int h(x)\pi(\mathrm{d}x)$$

is guaranteed by the LLN when the X_i's are iid from π and h is integrable against π and (b) the variations of $(\widehat{\mathfrak{I}}_n - \mathfrak{I})$ are asymptotically Gaussian if $\mathrm{var}_\pi(h(X))$ is finite. Therefore, if $\widehat{\sigma}_N^2$ is a convergent estimator of $\mathrm{var}_\pi(h(X))$, the CLT tells us that $\mathfrak{I}$ lies in the interval

$$[\widehat{\mathfrak{I}}_N - 2\widehat{\sigma}_N/\sqrt{N}, \widehat{\mathfrak{I}}_N + 2\widehat{\sigma}_N/\sqrt{N}],$$

with probability close to 0.95 for N large enough and the normal approximation implies that the pair $(\widehat{\mathfrak{I}}_N, \widehat{\sigma}_N)$ is sufficient for this assessment. Similar (albeit more sophisticated) results are available for the results of MCMC experiments.

Clearly the CLT is only informative about the variation of the approximation of $\mathfrak{I}$ by $\widehat{\mathfrak{I}}_N$ and does not tell us anything about on the probabilistic behaviour of the sequence $(\widehat{\mathfrak{I}}_n)_{1\leq n\leq N}$ and, in particular, on its expected variability as a sequence. In other words, the CLT is not helpful in assessing where $(\widehat{\mathfrak{I}}_n)_{1\leq n\leq N}$ could lie, were we to start a new simulation experiment.

As explained below, we need to know the (approximate) distribution of $(\widehat{\mathfrak{I}}_n)_{1\leq n\leq N}$ to gather information about its variability as a sequence without running an extensive Monte Carlo experiment that would require many replications of $(\widehat{\mathfrak{I}}_n)_{1\leq n\leq N}$. Such an approximation is found in Donsker's theorem (Section 4), and relates the sequence to a Brownian motion $\{W(t)\}_{0\leq t\leq 1}$.

3.2 Monte Carlo confidence bands

The primary purpose of a Monte Carlo experiment is to run a simulation algorithm till the approximation of $\mathfrak{I}$ by $\widehat{\mathfrak{I}}_N$ is satisfactory. However it can be the case that the variability of the sequence $(\widehat{\mathfrak{I}}_n)_{1\leq n\leq N}$ is of interest.

First, in MCMC experiments, we often use a crude graphical assessment of the stability of $(\widehat{\mathfrak{I}}_n)_{1\leq n\leq N}$ to make a preliminary decision about convergence to stationarity of the related Markov chain and may decide to increase the number of iterations or more radically to modify the MCMC kernel if this stability is not satisfactory, even though another run of the MCMC experiment, simply based on another sequence of pseudo-random seeds, could have produced a very different shape and led to a different conclusion, e.g. to continue the MCMC experiment and use more simulations. It thus seems necessary to provide a confidence assesment on the shape of an MCMC sequence in order to quantify this kind of decision.

Second, there exist many instances (see, e.g. Fishman, 1996) when a *control variate* is available, namely, when there exists a non-constant function h_0 such that

$$\mathfrak{I}_0 = \int h_0(x)\pi(\mathrm{d}x)$$

is known. When h and h_0 are similar, this control variate can be used to rank alternatives to iid simulation from π for the approximation of $\mathfrak{I}$. For instance, if importance sampling is used instead, with importance density q, the performance of the corresponding importance sampling estimate

$$\frac{1}{n}\sum_{i=1}^{n}\left\{h_0(Y_i)\frac{f_\pi(Y_i)}{q(Y_i)} - \mathfrak{I}_0\right\} \tag{6}$$

(where the Y_i's are an iid sample from the density $q(y)$, and $f_\pi(x)$ is the density of the distribution π) can be compared with iid sampling via the confidence band: if (6) escapes the confidence band derived from

$$\widehat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left\{h_0(X_i) - \mathfrak{I}_0\right\}^2,$$

where the X_i's are an iid sample from π, it is unlikely that its variance is smaller than the iid variance and therefore another importance density should be chosen without running the simulation experiment till the end.

Third, a similar (but more generic) use of the confidence band can be found in *calibration* experiments for importance sampling and MCMC kernels where the parameters governing the importance function or the Markov transition kernel are tailored towards more efficiency: once an initialising run has been produced, with a corresponding estimate $\widehat{\mathfrak{I}}_N^0$ of $\mathfrak{I}$, a first confidence band can be constructed and subsequent values of the parameters can be rejected as soon as the sequence

$$\frac{1}{n}\sum_{i=1}^{n}\left\{h(Y_i) - \widehat{\mathfrak{I}}_N^0\right\}$$

gets out of the band. If a new value of the parameters succeeds to stay within the band for N iterations, a new and more accurate estimate of $\mathfrak{I}$ can be constructed and a tighter confidence band used for the same purpose.

3.3 Sufficiency and sequential decision making

It follows from the previous section that the use of the confidence bands are mostly restricted to *sequential* settings. Indeed, if we wait till the Monte Carlo experiment is over, producing a sample $(X_1, \ldots, X_N)$ from π and approximating $\pi(h)$ by $\delta_N(1)$, the evaluation of this experiment will be based on the final output $\delta_N(1)$ and its empirical variance $\gamma_N(1)$ and not on its behaviour for $t \in [0,1]$, by a sufficiency argument (though that is strictly valid only in the normal case). In a static setting, the confidence bands for the sequence $\{\delta_N(t)\}_{t\in[0,1]}$ are therefore of little relevance and moreover they suffer from the major drawback that they do not preserve the exchangeability of the original sample $(X_1, \ldots, X_N)$: it may well happen that a sequence $\{\delta_N(t)\}_{t\in[0,1]}$ remains within the band while another ordering of the sample may produce a sequence that moves outside the band.

In sequential settings, the perspective is quite different however. We need to decide whether to abort a costly simulation experiment before its conclusion and therefore have to condition on the sequence of realised values *so far* whether or not they fit the ultimate confidence region. As explained above, this is pertinent in calibration experiments.

4 Brownian representations for Monte Carlo sequences

4.1 A functional CLT

The sequence of estimates $(\widehat{\mathfrak{I}}_n)_{1\leq n\leq N}$ is connected with the random step function

$$\{\delta_N(t)\}_{t\in[0,1]} = \left(\frac{1}{[Nt]_e} \sum_{i=1}^{[Nt]_e} h(X_i) \right)_{t\in[0,1]}$$

where $[x]_e$ denotes the integer part of x, that is, the greatest integer not exceeding x. By convention here and in the following, we take $\delta_N(t) = 0$ for $t < 1/N$. (Note that $\delta_N(1) = \widehat{\mathfrak{I}}_N$). The standard CLT can be formulated in terms of δ_N: for a fixed $t \in [0,1]$, if

$$0 < \mathbb{V}_\pi[h(X)] = \int h^2(x)\pi(\mathrm{d}x) - \pi(h)^2 < \infty\,,$$

the random variable

$$\sqrt{[Nt]_e} \left(\frac{\delta_N(t) - \pi(h)}{\mathbb{V}_\pi[h(X)]^{1/2}} \right)$$

converges in distribution to a standard $\mathscr{N}(0,1)$ random variable. In a standard Monte Carlo setting, it is also customary to replace $\mathbb{V}_\pi[h(X)]$ with an estimate. Introduce the random function

$$\{\gamma_N(t)\}_{t\in[0,1]} = \left(\frac{1}{[Nt]_e} \sum_{i=1}^{[Nt]_e} (h(X_i) - \delta_N(t))^2 \right)_{t\in[0,1]}\,.$$

For a fixed $t \in [0,1]$, if $0 < \mathbb{E}_\pi[|h(X)|^2] < \infty$ then $\gamma_N(t)$ almost surely converges to $\mathbb{V}_\pi[h(X)]$ and the random variable

$$\sqrt{[Nt]_e} \left(\frac{\delta_N(t) - \pi(h)}{\gamma_N(t)^{1/2}} \right)$$

also converges in distribution to a standard normal r.v. by virtue of Slutsky's theorem (Billingsley, 1995).

The functional extension of the CLT is called *Donsker's theorem*: it states that both the random functional

$$\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{[Nt]_e} \left(\frac{h(X_i) - \pi(h)}{\sqrt{\mathbb{V}_\pi[h(X)]}} \right) \right\}_{t\in[0,1]}$$

and its plug-in approximation

$$\left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{[Nt]_e} \{h(X_i) - \pi(h)\} \right)_{t\in[0,1]}$$

converge in distribution (when N goes to infinity) to a standard Brownian motion on $[0,1]$, denoted $\{W(t)\}_{t\in[0,1]}$. Donsker's theorem thus relates the Monte Carlo output with the confidence band constructed in Section 2 for the Brownian motion. In particular, we can readily see how to implement the control variate check introduced in the previous section. Based on the optimal bound $u^\star$ of (4), the region

$$\left\{ \omega\,;\, \omega(t) \in \left[\pi(h) - u^\star(t)\frac{\sqrt{N\gamma_N(1)}}{[Nt]_e}, \right. \right. \qquad (7)$$
$$\left. \left. \pi(h) + u^\star(t)\frac{\sqrt{N\gamma_N(1)}}{[Nt]_e} \right] , t \in [0,1] \right\}$$

contains $(\delta_N(t))_{t\in[0,1]}$ with asymptotic probability $1-\alpha$.

Nonetheless, this setting does not apply to the main bulk of simulation experiments since they are concerned with the approximation of an unknown

quantity $\pi(h)$ that cannot be used within the random functionals of Donsker's theorem or in (7). For the calibration use mentioned above, as well as for other applications, we thus need to study an alternative confidence band based on the replacement of $\pi(h)$ with the final estimate $\delta_N(1)$.

4.2 A generic confidence band

As pointed out previously, the quantity of interest from a Monte Carlo point of view is rather a confidence region that should contain a (random) Monte Carlo sequence with a given (asymptotic) probability. Formally, given an iid sample $Y_1, \ldots, Y_N$ from π that is independent of $X_1, \ldots, X_N$, associated with the estimators $\delta_N(1)$ and $\gamma_N(1)$, the decomposition

$$\left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{[Nt]_e} \{h(Y_i) - \delta_N(1)\}\right)_{t\in[0,1]} = \left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{[Nt]_e} \{h(Y_i) - \pi(h)\}\right)_{t\in[0,1]} - \left(\frac{[Nt]_e\, \gamma_N(1)^{-1/2}}{\sqrt{N}} \{\delta_N(1) - \pi(h)\}\right)_{t\in[0,1]}$$

implies that the scaled cumulative estimates converge in distribution to a continuous random process that is the sum of a standard Brownian motion $\{W(t)\}_{t\in[0,1]}$ and of a random linear function $(tU)_{t\in[0,1]}$, U being a standard normal random variable independent of $\{W(t)\}_{t\in[0,1]}$. This decomposition also applies when the Y_i's are generated from an importance density q and $h(Y_i)$ is replaced with $H(Y_i) = h(Y_i)f_\pi(Y_i)/q(Y_i)$, as well as in MCMC setups.

Therefore, if we set $Z(t) = W(t)+tU$, $t \in [0,1]$, we need to develop an equivalent of the confidence band of Section 2. In Section 5, we construct a function v on $[0,1]$ such that $(0 < \alpha < 1)$

$$\mathbb{P}(-v(t) \le Z(t) \le v(t), t \in [0,1]) = 1 - \alpha\,. \qquad (8)$$

This bound v can be used to assess the variability of a Monte Carlo sequence

$$\{\delta'_N(t)\}_{t\in[0,1]} = \left(\frac{1}{[Nt]_e} \sum_{i=1}^{[Nt]_e} H(Y_i)\right)_{t\in[0,1]}$$

with expectation $\pi(h)$ against the reference variability of the sequence $\{\delta_N(t)\}_{t\in[0,1]}$; namely the band

$$\left\{\omega\,; |\omega(t) - \delta_N(1)| \le v(t)\frac{\sqrt{N\gamma_N(1)}}{[Nt]_e}\,, t \in [0,1]\right\} \qquad (9)$$

has high probability of containing $\{\delta'_N(t)\}_{t\in[0,1]}$ if its variability is smaller than the reference variability. In a calibration experiment, this implies rejection of a simulation parameter at the point when the sequence $\{\delta'_N(t)\}_{t\in[0,1]}$ leaves the band.

As an aside, note that it is possible to use the process

$$\left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{[Nt]_e} \{h(Y_i) - h(X_i)\}\right)_{t\in[0,1]}\,,$$

since it converges in distribution to a continuous random process which is the sum of two independent standard Brownian motions or, equivalently, to the process $\{F(t)\}_{t\in[0,1]} = \{\sqrt{2}W(t)\}_{t\in[0,1]}$. However, the variability of this quantity is higher than when using the sufficient estimate $\delta_N(1)$ for the approximation of $\pi(h)$, and so there is thus little incentive in using this alternative.

5 Optimal confidence regions for a Brownian motion with random drift

The construction of confidence regions of level $1-\alpha$ is very similar to the work described above, with the difference that the limiting process is now $W(t)+tU$. The optimisation problem is thus

$$\min_{u\ge 0} \int_0^1 u(t)\mathrm{d}t$$

under the constraint that $P_2(u) = \mathbb{P}(-u(t) \le W(t) + tU \le u(t), t \in [0,1]) = 1-\alpha$.

Similar difficulties beset the derivation of the optimal bound and once more we use approximate boundaries to evaluate the boundary-crossing probability, derived either from the local time representation or by "crude" Monte Carlo methods.

5.1 Local time solution

The resolution of Section 2.1 also applies to this setting, in the sense that, for the modified problem (3),

we have an explicit solution given in (4). The difference with the above section is that, as we are dealing with $Z(t) = W(t) + tU$, we have $\tau(s) = s(1+s)$, and so the extremal envelopes are now given by

$$u^\star(s) = \psi(\kappa s(1+s))\sqrt{s(1+s)} \qquad (10)$$

if $\kappa s(1+s) \leq 1/\sqrt{e}$ and $u^\star(s) = 0$ otherwise, for varying κ.

Once again, use of *Mathematica* or *R* allows for the explicit computation of these optimal solutions $u^\star$. By way of example, Figure 2 shows upper and lower envelopes $\pm u^\star$ at $\kappa = 1/20$ together with 100 simulations of the Brownian motion with random drift produced from the empirical mean, $W(t) + tU$.

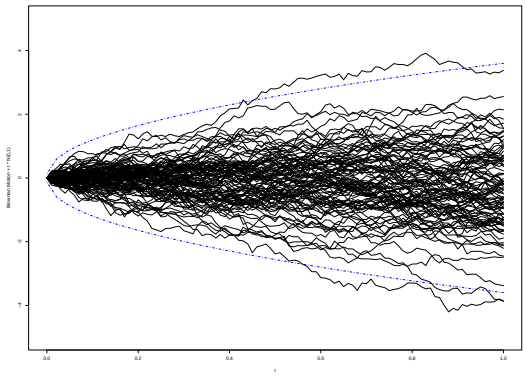

Figure 2: 100 Brownian with random drift trajectories together with a two-sided version of the extremal $u^\star$ for $\kappa = 1/20$.

For our determination of the $1-\alpha$ confidence region, we must thus calibrate κ to achieve a given coverage. Using the simulation experiment detailed below, we found $\kappa = 0.095$ for $\alpha = .05$.

5.2 Numerical results

Since we do not have a Volterra expression as for the Brownian motion, we use a cruder Monte Carlo approach to derive coverage probabilities and optimal values of parameterised boundaries.

We use a partition $t_0 = 0 < t_1 < t_2 < \ldots < t_n = 1$ of the interval $[0,1]$ of size $n \geq 1$, with $\beta_i = u(t_i)$. Obviously, if u is a continuous function on $[0,1]$, then

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=1}^{n} [-\beta_i \leq Z(t_i) \leq \beta_i]\right) = P_2(u)\,.$$

Class	$u(t)$	Minimum	$\int_0^1 u(t)\mathrm{d}t$
1	a	$a = 2.94$	2.94
2	$a + bt$	$a = 1.2$, $b = 2.1$	2.25
3	$a + b\sqrt{t}$	$a = 0.1$, $b = 3.15$	2.2
$\star$	$u^\star(t)$	$\kappa = .095$	2.189

Table 2: Optimal choices of parameters for specific bounds and coverage $1 - \alpha = .95$.

Moreover, $\mathbb{P}\left(\bigcap_{i=1}^{n} [-\beta_i \leq Z(t_i) \leq \beta_i]\right)$ can be estimated by a Monte Carlo approximation, namely

$$\hat{P}_2(u)_{n,N} = \frac{1}{N}\sum_{j=1}^{N}\prod_{i=0}^{n} \mathbb{I}_{[-\beta_i \leq Z_j(t_i) \leq \beta_i]}$$

where $Z_1, \ldots, Z_N$ are N independent standard Brownian motions on $[0,1]$ with random drift. As N goes to infinity, $\hat{P}_2(u)_{n,N}$ converges almost surely to

$$\mathbb{P}\left(\bigcap_{i=1}^{n} [-\beta_i \leq Z(t_i) \leq \beta_i]\right)$$

and, as n and N both go to infinity, $\hat{P}_2(u)_{n,N}$ is a convergent estimator of $P_2(u)$.

In our numerical experiment, we chose $N = 1,000,000$ and $n = 1000$. For the same three parameterised classes of functions $u(\cdot)$ as above, we have evaluated the function u that minimises its subgraph under the coverage constraint $P_2(u) = 0.95$. Table 2 presents the results, along with the corresponding local time solution $u^\star$. Figure 3 plots $5,000$ paths $W(t) + tU$ together with the four functions.

Once again, the gain brought by the local time solution or by other classes of boundary functions, like $u(t) = a + bt + ct^2 + d\log(t)$, is reassuringly minimal when compared with the simple approximation $u_0(t) = 0.1 + 3.15\sqrt{t}$. In practice, since u_0 is almost indistinguishable from $u^\star$, it seems more convenient to use u_0 for the evaluation of Monte Carlo variability.

6 Example

We consider a toy problem when π is the $\mathcal{N}(0,1)$ distribution and $h(x) = \exp(x^2/4.01)$. This specific function h is chosen because it is near to the borderline for an infinite variance estimator—had we chosen

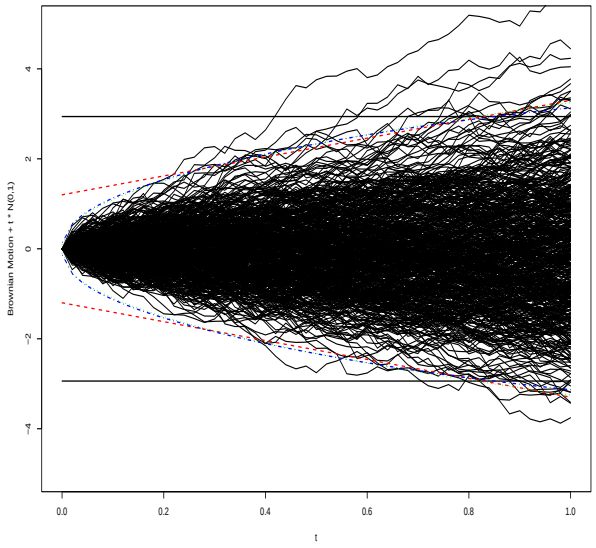

Figure 3: $5,000$ paths of the process $(Z(t))_{t\in[0,1]}$ and the four boundaries of Table 2.

4.0 instead of 4.01 in the denominator, the regular Monte Carlo estimator would have had an infinite variance and the CLT would not then apply. The Monte Carlo estimator of $\pi(h)$ is

$$\delta_N(1) = \frac{1}{N}\sum_{i=1}^{N} \exp\left(X_i^2/4.01\right)$$

where $X_1, \ldots, X_N$ is an iid $\mathcal{N}(0,1)$ sample. The average behaviour of the sequence of $\delta_N(t)$'s is illustrated by Figure 4, for 1000 replicas of $N = 1000$ simulations. Note the interesting feature that one replica makes a huge jump around $t = 0.55$: this is typical for estimators with infinite (or almost infinite) variances but this type of jump only occurred once over the 1000 replicas. This means not only that a predictive confidence band should not include this extreme behaviour, but also that were we to run a single simulation and observe this type of jump then the resulting confidence band would be much larger. The confidence band, as derived in Section 5 using (9), is represented in Figure 5 for one Monte Carlo sequence. This band is centered at the final value of the estimate $\delta_N(1)$ and does not present irregular jumps on its boundary. (In that case, the band contains the simulated sequence although this is not always the case. For instance, the special replica that takes a huge jump around $t = 0.55$ would escape its own confidence band.)

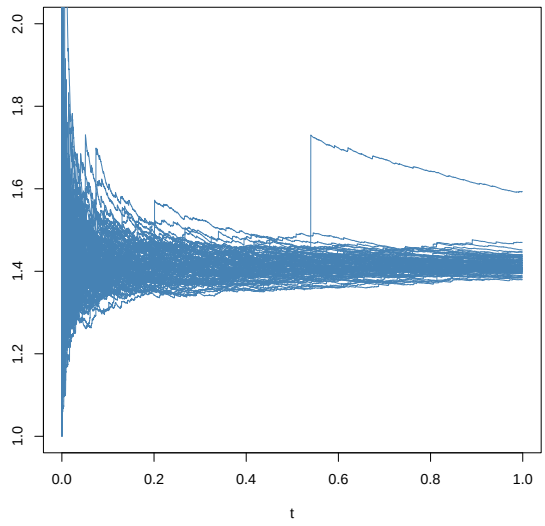

Figure 4: 1000 replicas of a sequence of Monte Carlo estimators $\delta_N(t)$ for $N = 1000$.

For illustrative purposes, we now consider a simulation design experiment with the same function h and the same distribution π. If we resort to importance sampling from $\mathcal{N}(0,\tau^2)$ distributions, τ^2 being a simulation parameter requiring calibration, we can produce the alternative estimator of $\pi(h)$ given by

$$\delta'_N(1) = \frac{1}{N}\sum_{i=1}^{N} \exp\left(Y_i^2/4.01\right) \times \frac{\exp\left(-Y_i^2/2\right)}{\tau^{-1}\exp\left(-Y_i^2/2\tau^2\right)}\,,$$

where $h(Y_i)$ is multiplied by the importance weight. Starting with $\tau^2 = 1$ and increasing it by steps, we thus build a sequence of confidence bands and of simulated sequences $(\delta'_N(1))_{0\le t\le 1}$ till the current sequence escapes the previous confidence band. Figure 6 shows the result of this experiment: while $\tau^2 < 2$, the sequences remain within the earlier confidence band while the estimated variance and corresponding confidence band decrease. As soon as $\tau^2 > 2$, the sequence $(\delta'_N(1))_{0\le t\le 1}$ escapes the band rapidly, which shows that further increases of τ^2 are counterproductive. This coincides with the theoretical assessment of the problem since the optimal value of τ^2 is

$$\tau^2_\star = \frac{1}{1-2/4.01} = 1.995025\,,$$

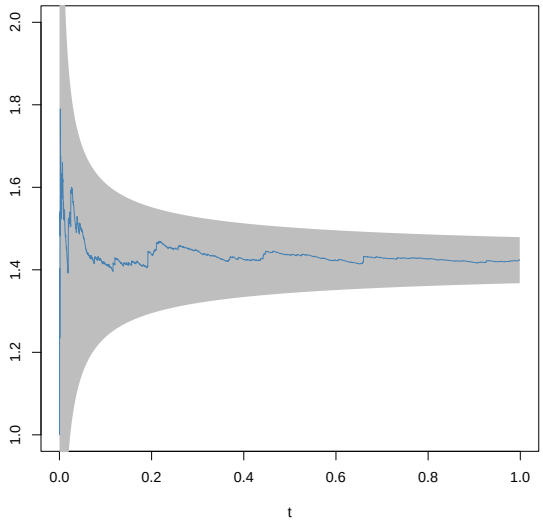

Figure 5: Confidence band at level 0.95 derived from (9) for one realisation taken from Figure 4.

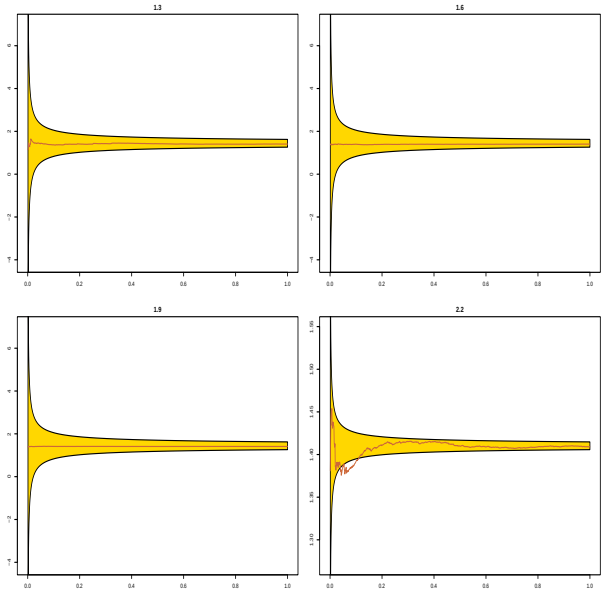

Figure 6: Sequence of confidence bands for increasing values of τ^2 ($\tau^2 = 1.3, 1.6, 1.9, 2.2$) and $N = 1000$. Note the estimate is nearly of zero-variance for $\tau^2 = 1.6$, 1.9.

which produces a zero variance estimator. So this toy problem exemplifies how the confidence band can be used sequentially in realistic and expensive simulation studies.

7 Conclusion

The bands constructed in this paper are universal in that they can be used in any Monte Carlo experiment, given that the bound u has only to be determined once for a confidence level α. We also stress that the case of Markov Chain Monte Carlo settings causes no special difficulty provided only that the correct functional CLT is applied.

Acknowledgements

The first author is grateful to Sigurd Assing and to Larbi Alili for helpful conversations about boundary problems. The second author is grateful to Catherine Loader for sending him *R* programs related to the computation of the boundary crossing probabilities, as well as for useful discussions. The third author is grateful to both the Institute for Mathematical Sciences, National University of Singapore, for inviting him to attend the workshop on Markov Chain Monte Carlo in March 2004, thus giving him the opportunity to collaborate with the first author, and to George Casella for discussions about a preliminary perspective on that problem that eventually led to Chapter 4 in Robert and Casella (2004). Feedback from the audience of the ISBA 2004 poster session in Viña del Mar, Chile, is also gratefully acknowledged. A conversation with Marc Yor in the premises of this research was equally helpful. Finally, comments from two anonymous referees were very helpful in improving the clarity of the paper.

References

Anderson, T. (1960). A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size. *Ann. Math. Statist.*, 31(1):165–197.

Billingsley, P. (1995). *Probability and Measure*. John Wiley, New York, third edition.

Borodin, A. and Salminen, P. (2002). *Handbook of Brownian motion - facts and formulae*. Birkhauser Verlag, 2 edition.

Corless, R., Gonnet, G., Hare, E., Jeffrey, D., and Knuth,

D. (1996). On the Lambert W function. *Adv. Computational Math.*, 5:329–359.

Daniels, H. (1996). Approximating the first crossing-time density for a curved boundary. *Bernoulli*, 2:133–143.

Durbin, J. (1971). Boundary crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. *J. Appl. Prob.*, 8:431–453.

Durbin, J. (1992). The first-passage density of the Brownian motion process to a curved boundary. *J. Appl. Prob.*, 29:291–304.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, volume 2. John Wiley, 2 edition.

Fishman, G. S. (1996). *Monte Carlo*. Springer Series in Operations Research. Springer-Verlag, New York. Concepts, algorithms, and applications.

Hall, W. (1997). The distribution of Brownian motion on linear stopping boundaries. *Sequential Anal.*, 4:345–352.

Hollander, M. and Wolfe, D. A. (1999). *Nonparametric statistical methods*. Wiley Series in Probability and Statistics: Texts and References Section. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Lerche, H. R. (1986). *Boundary Crossing of Brownian Motion*. Lecture Notes in Statistics. Springer-Verlag.

Li, W. V. (2003). The first exit time of a Brownian motion from an unbounded convex domain. *Ann. Prob.*, 31(2):1078–1096.

Loader, C. and Deely, J. (1987). Computations of Boundary Crossing Probabilities for the Wiener Process. *J. Stat. Comp. Simul.*, 27:96–105.

Novikov, A., Frishling, V., and Kordzakhia, N. (1999). Approximation of boundary crossing probabilities for a Brownian motion. *J. Appl. Prob.*, 36:1019–1030.

Park, C. and Schuurmann, F. (1976). Evaluation of barrier crossing probabilities for Wiener path. *J. Appl. Prob.*, 17:197–201.

Potzelberger, K. and Wang, L. (2001). Boundary crossing probability for Brownian motion. *J. Appl. Prob.*, 38:152–164.

Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann. Math. Statist.*, 41:1410–1429.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, second edition.

Wang, L. and Potzelberger, K. (1997). Boundary crossing probability for Brownian motion and general boundaries. *J. Appl. Prob.*, 34:54–65.